# Geometric Patterns: Algorithms and Applications

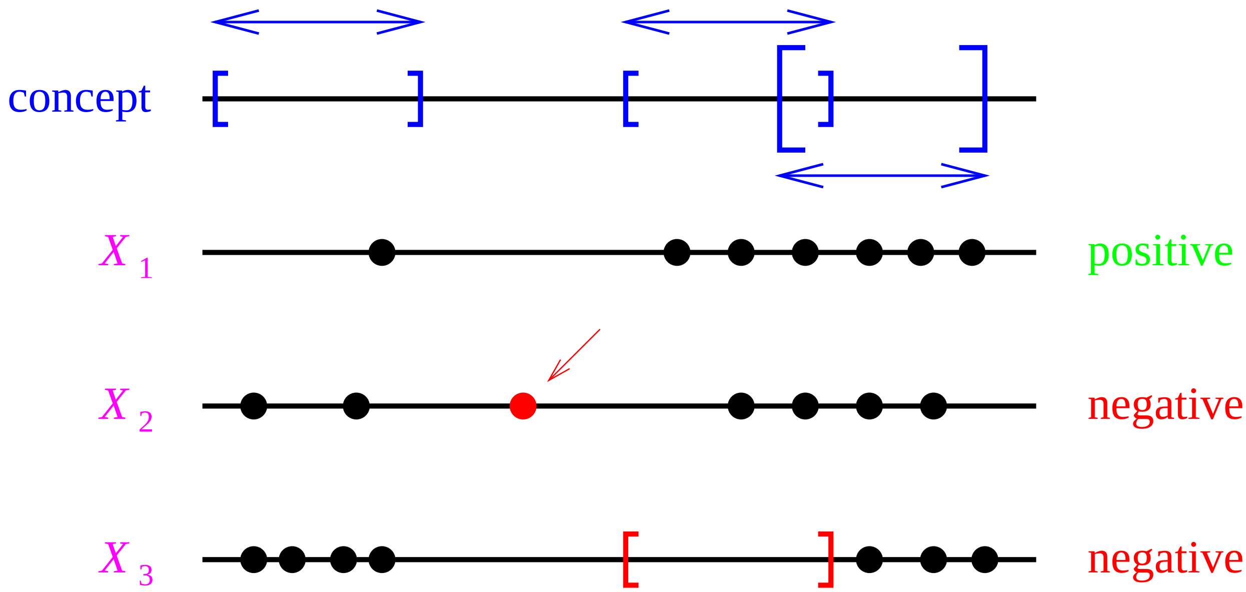
Stephen D. Scott
Dept. of Computer Science
University of Nebraska

July 2, 2000

# Concept Class of One-Dimensional Patterns

- Each concept $c$ is a set of fixed-width intervals on real line

- Each example $X$ is a set of points on real line




- Set of positives is set of patts within const dist of each other under Hausdorff metric:

$$\max \left\{ \max_{p \in P} \left\{ \min_{q \in Q} \{ dist(p,q) \} \right\}, \max_{q \in Q} \left\{ \min_{p \in P} \{ dist(p,q) \} \right\} \right\}$$

## Algorithm 1 (Batch)
[*Machine Learning* 96]

- Finds set of intervals consistent with positives, then performs greedy covering of negatives

- PAC algorithm: Given any $0 < \epsilon, \delta < 1$, can provably get

$$\Pr[\text{prediction error} \leq \epsilon] \geq 1 - \delta$$

  in time & examples $poly(1/\epsilon, 1/\delta, \text{ex. size}, \text{concept size})$

- Performed well on simulated test data

- Intolerant of noise (sensitive to individual training labels)

## Algorithm 2 (Batch)
[*Machine Learning* 99]

- Doesn't depend on individual examples, instead uses statistical queries [e.g. Kearns 93]

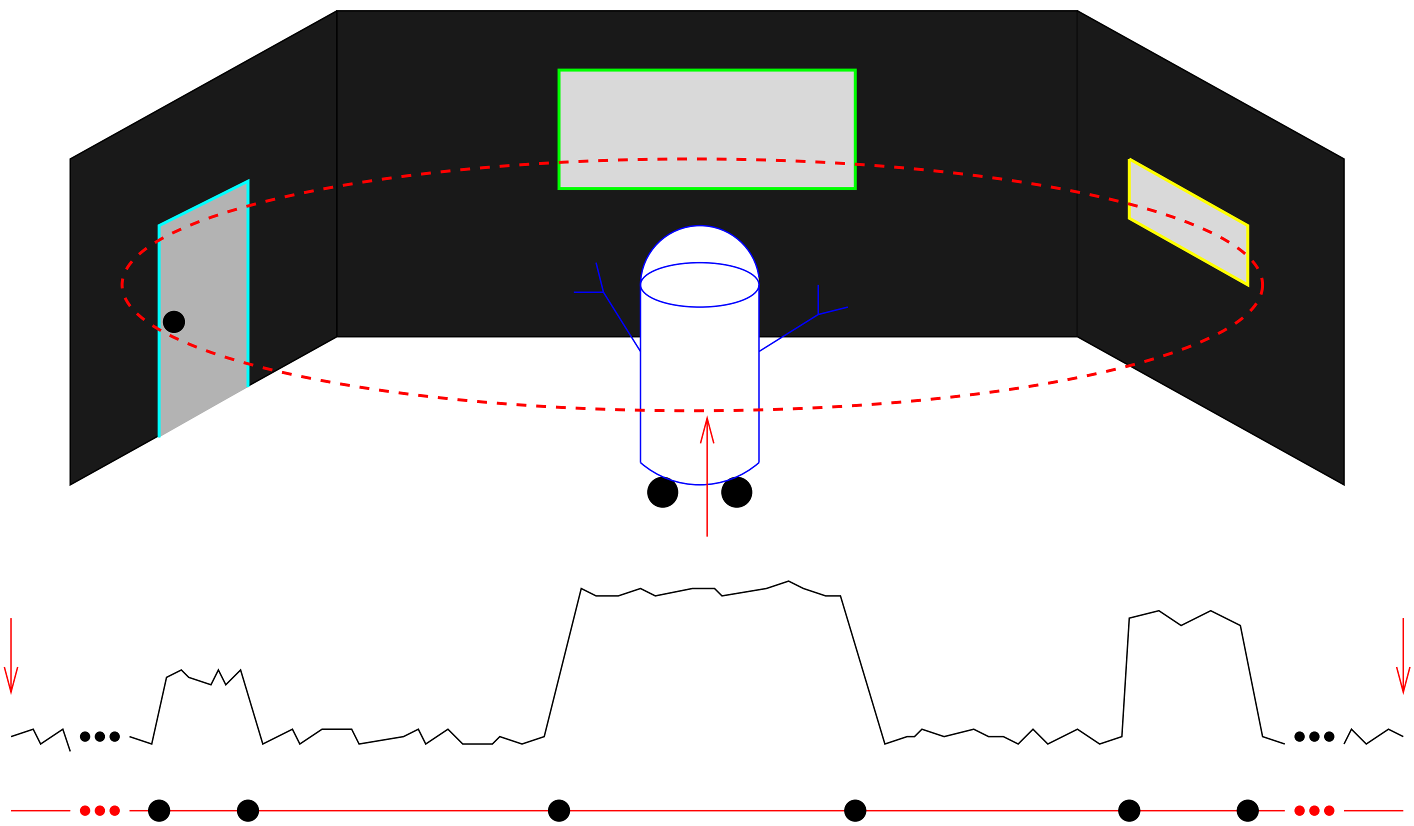
  "What's the probability that a random ex. is positive and has a point in $[a, b]$?"

- Uses SQs to greedily cover (reduce) false pos and false neg error

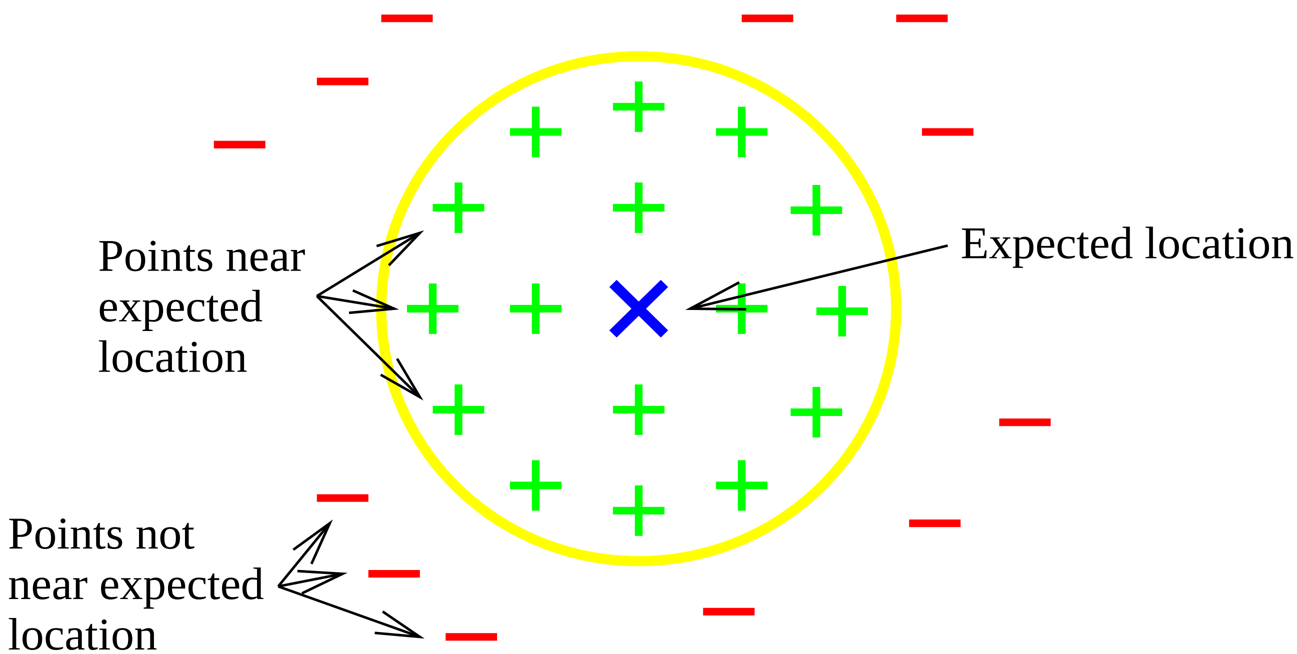
- Is a PAC algorithm that provably tolerates classification noise and other types of noise

- Performed well on noise-free and noisy simulated test data and OK on real data

# Mapping Robot Vision Data to 1D Patterns

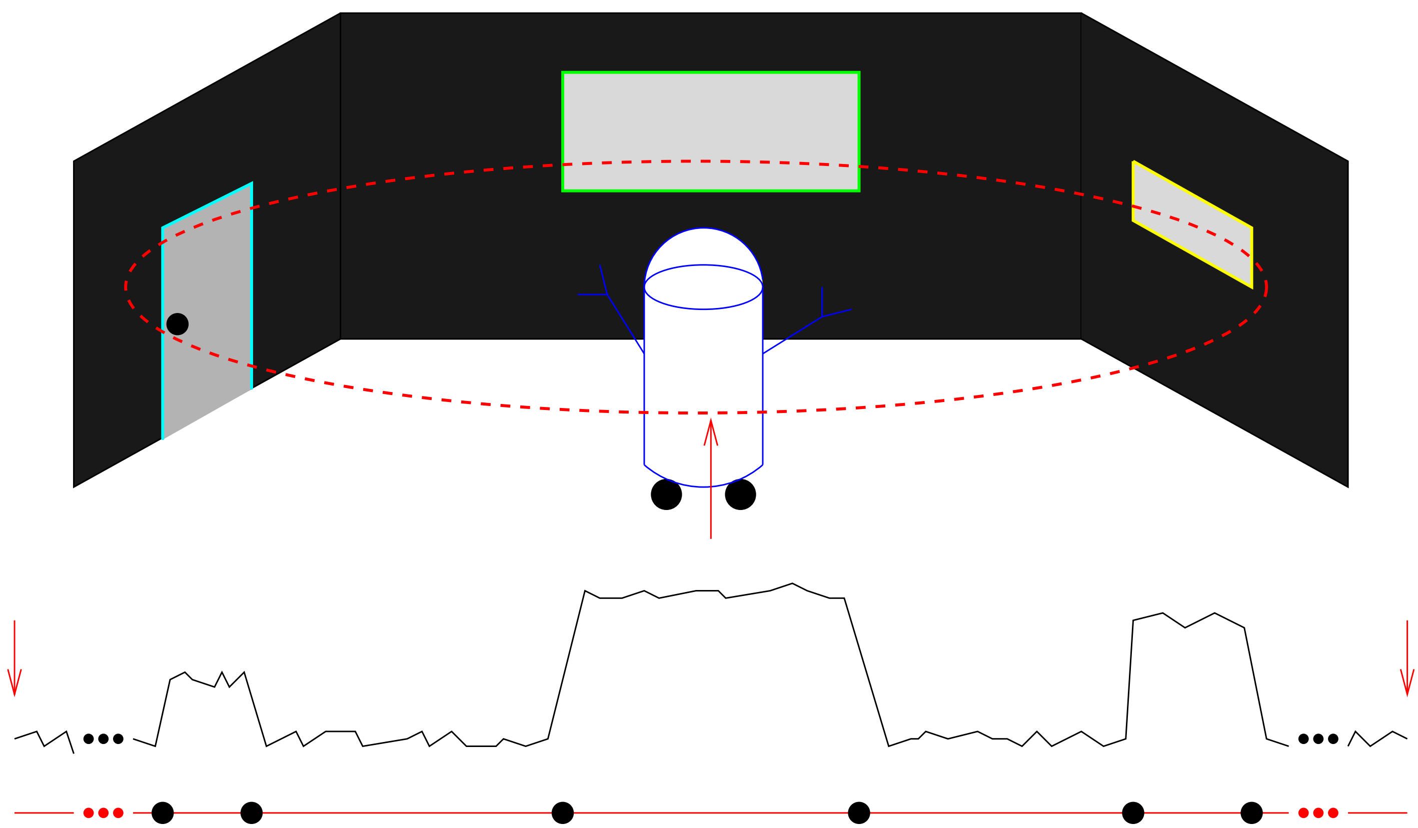


# Applying Learning

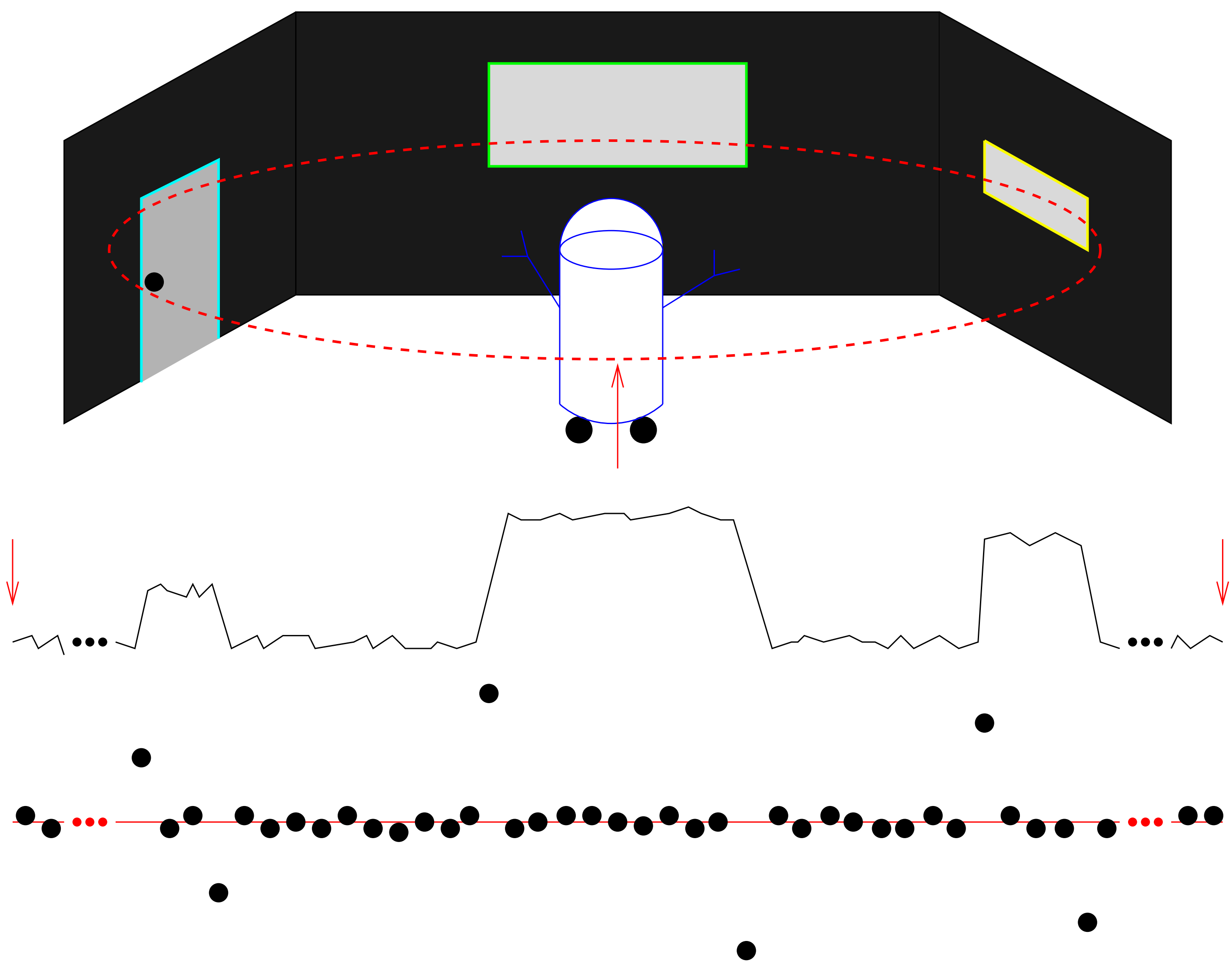


# Problem with 1D Mapping

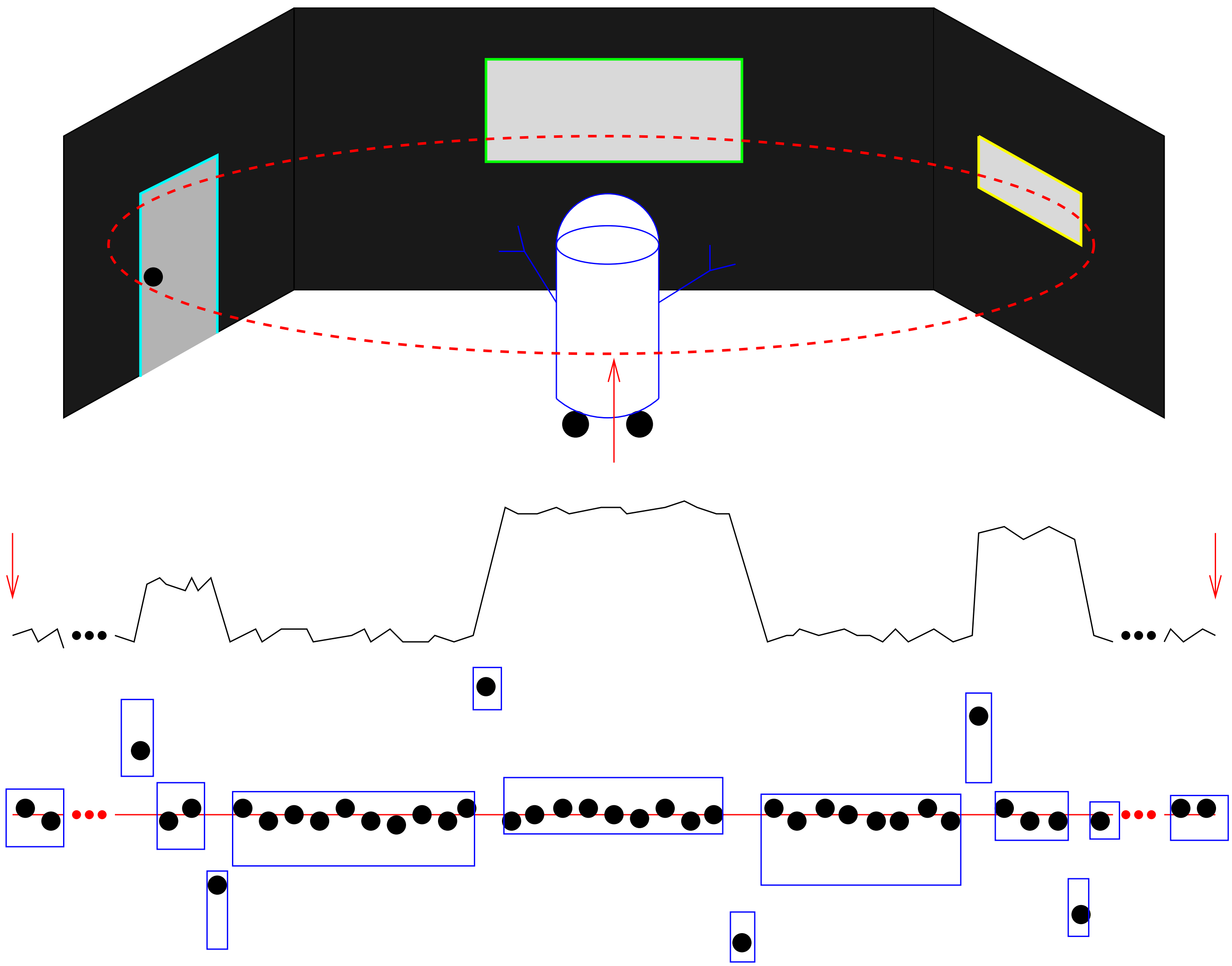
- Empirical work revealed that too many dissimilar 1D images map to similar patts



# New Mapping

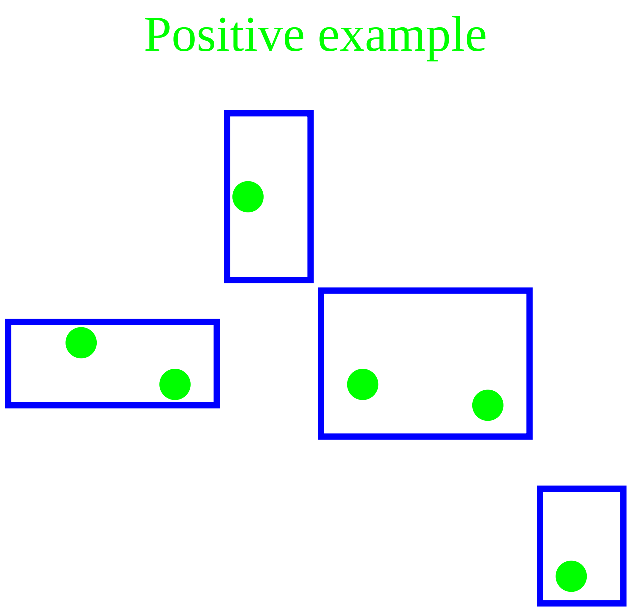


# New Mapping



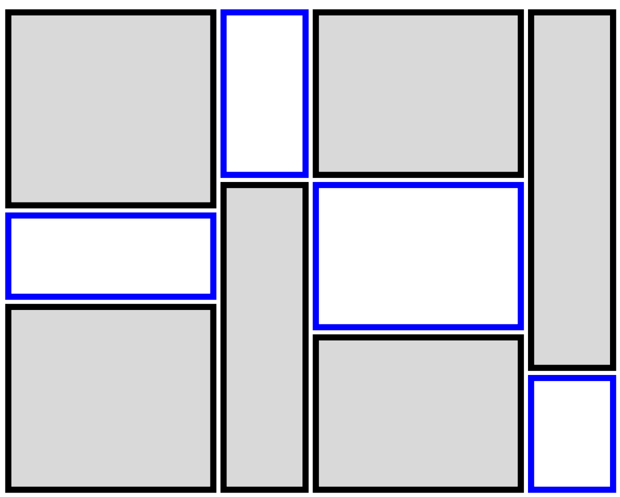
# New Concept Class

- $c =$ set of arbitrary, axis-parallel boxes




- Generalizes Hausdorff metric under weighted $L_\infty$ norm (for each box, each dimension can have different weight)

# New Concept Class



Example negative iff

A blue (unshaded) box is empty

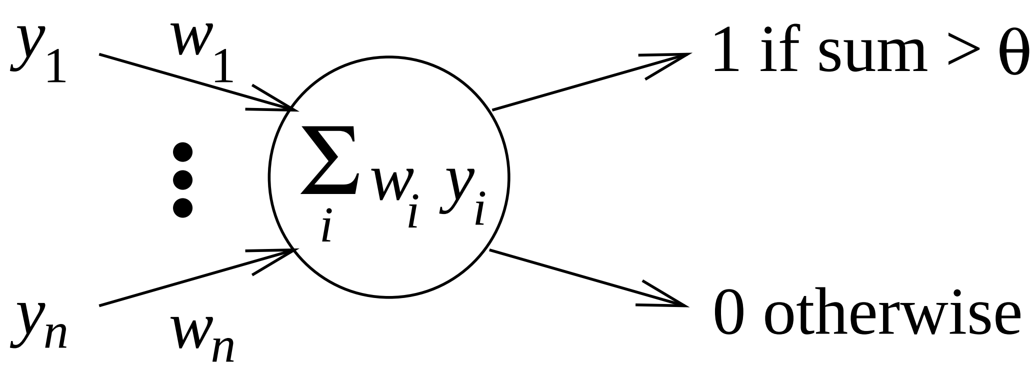
OR

A shaded box contains a point

- Classification representable as disjunction of boolean attributes

# Winnow [Littlestone 88]

$Y = y_1 \bullet\bullet\bullet\, y_n \qquad y_i \in \{0,1\}$




| pred | correct | name | update |
|---|---|---|---|
| 1 | 0 | demotion | if $y_i = 1$, $w_i = w_i/\alpha$ |
| 0 | 1 | promotion | if $y_i = 1$, $w_i = w_i \cdot \alpha$ |

- Can learn disjunction of $\leq k$ attributes with $O(k \log n)$ mistakes

## Algorithm 3 (On-line)

[COLT 97; *JCSS*, to appear]

- Reduce learning of concept to learning disjunction and apply Winnow

- Agnostic: Error guaranteed to be within a factor of $M_{opt}$ = best possible with our hypothesis class

  - If discrete space is $r \times s$, then mistake bound $O\left((k_{blue} + k_{shaded}) \cdot (M_{opt} + \log r + \log s)\right)$ and $k_{shaded} = O(k_{blue}^d)$

- Time is $poly(\log r, \log s, M_{opt}, n, k_{blue}^d)$

- Tolerant of concept shift [Auer & Warmuth 95]

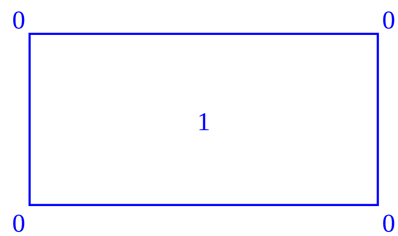
- Can map to PAC

- Can scale to arbitrary (fixed) dimension $d$, so can apply to 2D images + auxiliary information

# Fuzzy Patterns

- Labels and predictions now come from $[0,1]$ to allow for uncertainty (e.g. dist from landmark)

- Each fuzzy box $c$ has a membership function $\mu$, e.g. 1 at center, linear drop to 0 at corners:




- Combine boxes with aggregation operator:

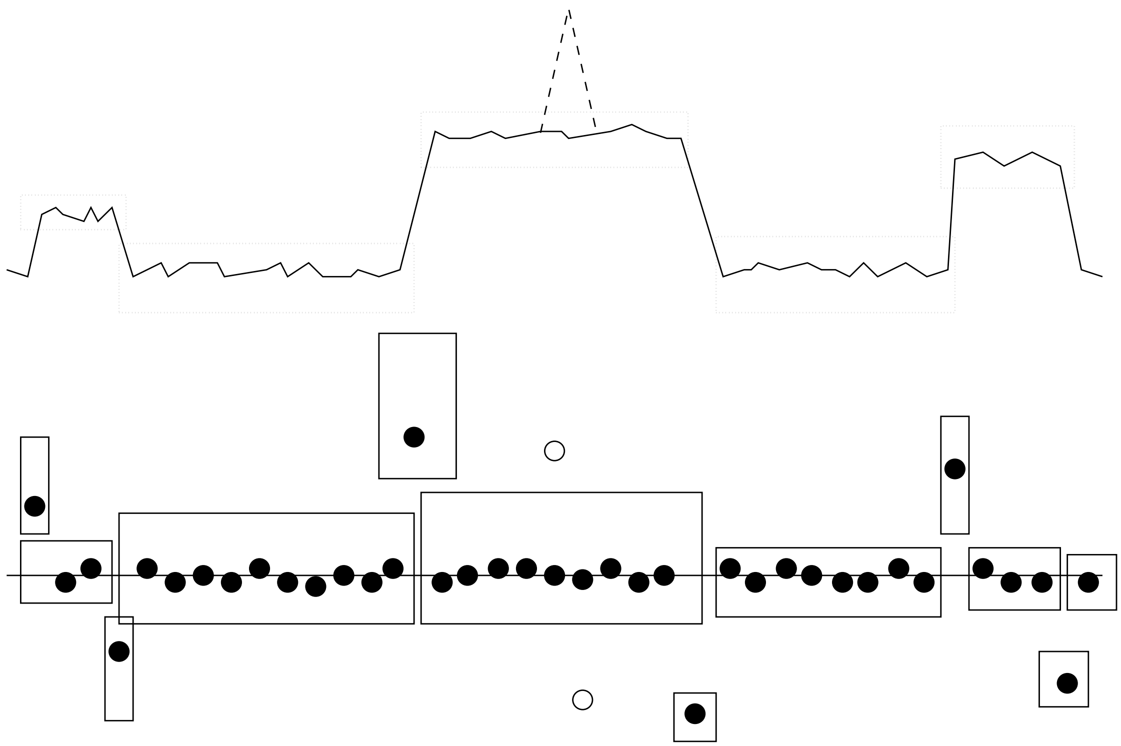
  – max-max: $\max_{c \in C}\{\max_{p \in P}\{\mu_c(p)\}\}$
  (Like unions of boxes)

  – minimax: $\min\left\{\min_{c \in C}\{\max_{p \in P}\{\mu_c(p)\}\}, \min_{p \in P}\{\max_{c \in C}\{\mu_c(p)\}\}\right\}$
  (Like fuzzy patterns)

  – average: $(1/m)\sum_{i=1}^{m} \max_{c \in C}\{\mu_c(p_i)\}$
  (Like fuzzy patts with noise tolerance)

# Fuzzy Patterns

Average Aggregation Operator



- Images (and patterns) with and without dashed spike very similar, but minimax labels first as near 1, second as 0

- Average labels first as near 1, second as a little less

## Algorithm 4

[In preparation]

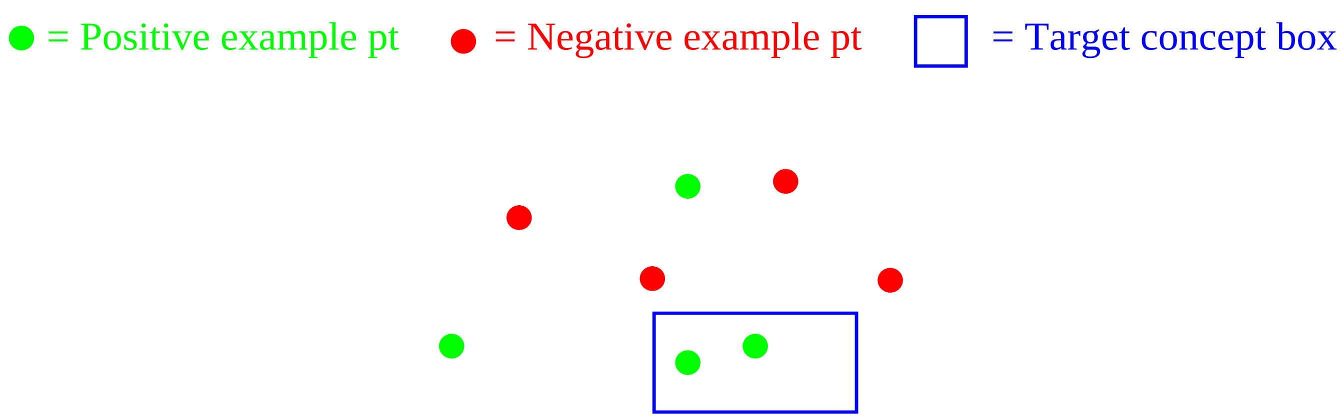
- Similar to algorithm 3, but use exponentiated gradient (EG) [Kivinen & Warmuth, 97] rather than Winnow

- Bookkeeping significantly more complex, but still polynomial for fixed $d$

- Loss bounds still agnostic, i.e. guaranteed not much worse than best collection of fuzzy boxes using same aggregation operator

## Other Applications

- Learning with multiple-instance exs
  - Drug activity prediction [e.g. Dietterich et al. 97]: each pt is shape of a molecule's conformation, concept box is shape constraints on at least one conformation for it to bind at a particular site
  - Content-based image retrieval [e.g. Maron & Ratan 98]: each point is some features of an image, concept box represents a desired set of features

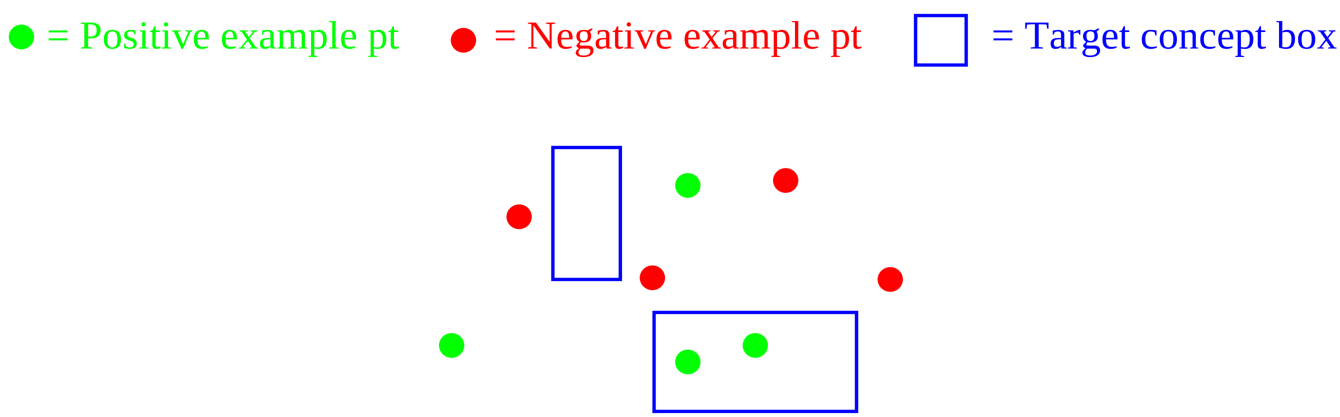



- Positive iff $\geq 1$ of its points in target box

## Other Applications
(cont'd)

- Unions of boxes with multiple-instance exs

- Positive iff $\geq 1$ of its points in some box




- Other conjunctive rules for positive classification possible, e.g. every box contains some point (can do e.g. antagonist drugs, where molecule must bind at multiple sites)

- Our algs can learn these & other variations

- Drawback: time complexity exponential in $d$

# Current Work

- Robot vision
  - Mappings from 1D, 2D data to patts
  - Evaluating Algorithm 3 to learn concepts
- Content-based image retrieval
  - Use our (and others') algorithms for multi-instance learning on already tested features, e.g. [Maron & Ratan 98]'s "blobs"
  - Other simple features, e.g. Fourier coefficients of shapes, elongatedness
  - Utilize the abilities of our algorithms to learn conjunctive multi-instance concepts
- Drug activity prediction (esp. antagonist drugs)
  - Must work around exponential time complexity (in $d$)